Target Language-Aware Constrained Inference for Cross-lingual Dependency Parsing
Tao Meng¹, Nanyun Peng², Kai-Wei Chang¹
¹University of California, Los Angeles ²University of Southern California

Overview

Task: Cross-lingual Dependency Parsing

We are trying to capture differences between languages.

Motivations

- Prior work: focus on capturing commonalities between languages.
- Leverage linguistic properties of the target to facilitate the transfer.

Contributions

- We explore corpus linguistic statistics derived from WALS features and compile them into corpus-wise constraints to guide the inference process during the test time.
- We improve the performances on 17 out of 19 target languages.

Background

Graph-Based Parser:

- Assigns a score for every word pair and conducts inference to derive a directed spanning tree with the highest accumulated score.
- Integer linear program (ILP) Inference: \( \max \sum y_{i,j} S_{ij}^{(k)} y_{k}(i,j) \)

Corpus-Statistics Constraints

Unary constraints:

- Statistics regarding a particular POS tag (POS).
  - E.g. Spanish:

  DET NOUN VERB DET NOUN ADP DET NOUN ADP NOUN PUNCT
  Este triunfo supuso su comienzo en el mundo de moda.

  Heads of NOUN appears on the left 82.9% of the time.

Binary constraints:

- Statistics regarding a pair of POS tags (POS₁, POS₂).
  - E.g. In Hindi, ADP appears on the right of NOUN in ADP-NOUN arcs 99.9% of the time

Inference with Corpus-Statistics Constraints

- Lagrangian Relaxation (Right).
  - Constrained inference problem can be formulated as an ILP:
    \( \max \sum y_{i,j} S_{ij}^{(k)} y_{k}(i,j) \) s.t. \( r_i - \theta_i \leq R(C, Y) \leq r_i + \theta_i, \ i \in [N] \)
  - Solve approximately by Lagrangian Relaxation:
    - Lagrangian multipliers \( \lambda \rightarrow \) relax the constraints.
    - Iteratively \( \lambda(t) \rightarrow \lambda(t+1) \)

- Posterior Regularization (Middle).
  - Treat the model as a probability model \( p_{θ} \):
    \( p_k(i,j) \propto \exp S_{ij}^{(k)} \)
  - Define the feasible set \( Q \) by constraints:
    \( r_i - \theta_i \leq R(C, q) \leq r_i + \theta_i, \ i \in [N] \)
  - Find the closest distribution in \( Q \) from \( p_{θ} \):
    \( q^* = \arg\min_{q \in Q} KL(q||p_{θ}) \)
  - MAP inference based on the feasible distribution \( q^* \):
    \( Y = \arg\max_{q \in Q} q^*(Y) \)

Results

- Significant improvements in low-resource languages. Keep or slightly improve the performance in common languages.

- Analysis about individual constraints and the relation between improvements and ratio gap (Highly related, Pearson 0.938).

Conclusion

- Improve 15 and 17 languages out of 19 with LR and PR, respectively.
- Languages with different word order from English improve significantly.
- Lagrangian relaxation has a greater average improvement, while posterior regularization improves more languages.
- Code and models:
  https://github.com/MtSomeThree/CrossLingualDependencyParsing/